International Journal of Computer Science Engineering
and Information Technology Research (IJCSEITR)
ISSN(P): 2249-6831; ISSN(E): 2249-7943
Vol. 4, Issue 4, Aug 2014, 59-68
© TJPRC Pvt. Ltd.

# A NEW APPROACH TOWARDS INFORMATION SECURITY BASED ON DNA CRYPTOGRAPHY

**ABHISHEK MAJUMDAR & MEENAKSHI SHARMA**

Department of Computer Science and Engineering, SSCET, Badhani, Punjab, India

## ABSTRACT

Information security plays a vital role in this era. During the transmission of data different sorts of attacks may happen and effects on data. To ensure the confidentiality and authenticity of the secret information a lot of cryptographic approaches have developed and researchers are still working on it to provide better approach towards information security. Cryptography can be defined as a process of transforming the sender's message to a secret format that can only be understood by the intended receiver. The DNA cryptography is a new area to achieve higher level of information security, where different features of the actual human DNA are followed. In this paper an algorithm is proposed where a long and strong 256-bit key is generated and use for the round encryption. Moreover a better level of encryption technique has been carried out. For this, four round operations are performed between the plain text and the generated key. Later on the resultant cipher text is transformed to a DNA sequence and appends some extra information bits within that to provide better security in the message against the intruders attack.

**KEYWORDS:** Plain Text, Key, DNA Sequence, Nucleotide, Intermediate Cipher Text, Hash Mapping

## INTRODUCTION

During the message transmission the main focus is on the security of the information. Due to this the concept of cryptography was introduced which means to convert a known text in any coded format which will not be revealed by any other third party except the intended sender and the receiver. Several cryptography approaches are proposed by a lot of researchers for several years. In their approaches they performed a number of complex mathematical computations to enhance the security of the information. In order to enhance data security and make the data more confidential effective encryption algorithms are required.DNA based encryption method is one of the recent technique in cryptographic field. Today, DNA based cryptography is taken as a most promising area of research by several researchers due to having the complex structural features and several special characteristics of the DNA. Out of them some used DNA computing, while some other incorporated biological properties of DNA strands and DNA sequence in their algorithms.In this paper a DNA based cryptographic approach is proposed where the message integrity is also maintained side by side with the information security. A 256 bit key is shared among the sender and the receiver and is used for the secret key generation that will be used the encryption phase. This secret key works on each plaintext blocks in four consecutive rounds by following an ordered manner and generate the cipher text and later on it will also converted to a DNA sequence to make it more secure and not understandable. Moreover to ensure the integrity of the message and the data origin authentication at the receiver end the message authentication code is used with a shared hash function.

## BIO-LOGICAL IDEA OF DNA

The DNA actually stands for Deoxyribo Nucleic Acid. In human body each cell contains a nucleus which characterizes all the physical and behavioral features of human body. They are packed into chromosomes. A DNA is nothing but a double helix made up of two strands where each strand can have either a purine or a pyramidine base. The purine bases are adenine (A) and guanine (G), while the pyrimidines bases are thymine (T) and cytosine (C), also known as the 4 basic nucleotide bases of DNA. In a double helix DNA the two strands are linked together where bases are bonded each other by hydrogen bonds: A with T and C with G, which is called the complementary pairs of DNA strands.

But to enhance the security and to increase the complexity of information this natural complementary rules can be changed. In a DNA sequence every three adjacent nucleotide bases forms a codon which maps to a unique amino acid that is used in protein synthesis. Another fact about is the primer which is nothing but a DNA sequence which is appended on both sides of the original DNA sequence that makes difficult to the intruder to identify the original DNA sequence.

## LITERATURE SURVEY

The research on DNA cryptography is still in a preliminary stage and still it needs a better level of security than the available ones. Lots of DNA cryptography methods are implemented, where several kind of encryption process was mentioned and a number of ways to exchange the secret keys was given. H. Z. Hsu, R. C. T. Lee et al. have proposed DNA cryptography approach in which they had pointed some interesting properties of DNA sequences to encrypt data. They discussed three methods, and for each method, they secretly select a reference DNA sequence [1]. Amal Khalifa and Ahmed Atito et al. discussed a method of text hiding where the text is encrypted using amino acid and DNA based playfair cipher and also use complementary rules to hide the resultant cipher text in a DNA sequence [2]. Mohammad Reza Abbasy, Pourya Nikfard et al. has given a DNA based cryptographic approach in which a sort of indexing method over the complementary DNA sequence was used [3]. Sabari Pramanik, Sanjit Kumar Setua et al. used a single stranded DNA string as the secret key whose length depends on the plain text and they used it to encrypt the plain text. Moreover they send the plain text as several DNA plain text packets by attaching the packet sequence number with each packet[4]. Suman Chakraborty, Sudipta Roy et al. had incorporated an idea of DNA based image encryption using soduko solution matrix to perform some computations on behalf of the message [5]. Nirmalya Kar, Atanu Majumder et al. had proposed a more secure and reliable encryption scheme by using the technologies of DNA sequence and DNA synthesis. They used three keys for encrypting the message along with a new method of key generation and key sharing. Instead of directly sharing the key, a session key holding the information regarding the actual encryption key was shared among two parties [6]. The researchers are implementing to enhance the security of the cipher text by appending extra coded information with it at different location of the cipher text [8]. Bibhash Roy, Atanu Majumder et al.has derived an encryption method in which two levels of encryption took place that was concerned with how DNA sequencing can be used in the field of cryptography [7]. Some of the researchers like Xing Wang, Qiang Zhang et al. derived a new way to show how cryptography works with DNA computing, it can transmit message securely and effectively. They have used RSA algorithm along with DNA computing theory [9]. Guangzhao Cui, Limin Qin et al. have worked on DNA molecules that are being explored for computing, data storage and cryptography. The encryption scheme here was based on the technologies of DNA synthesis, PCR amplification and DNA digital coding as well as the theory of traditional cryptography [10].

## PROPOSED METHOD

The Encryption method used here is somehow different from the other encryption algorithms till now implemented. In the earlier encryption schemes very basic concepts regarding the DNA were used such as simple XOR operation, primer addition, using complementary rules of DNA etc. But here the proposed DNA based encryption algorithm is little different from others. The previously implemented cryptographic algorithms provide only secrecy or confidentiality, but not the integrity which must be needed in some cases. Integrity can be defined as an information need which ensures the protection of the information from an unauthorized change. The proposed method provides a secured and reliable data transmission as well as message integrity. Here the overall method is done by 3 sub phases; these are the key selection and generation, DNA data encryption and the use of message detection code (MDC) to provide data integrity.

### Key Generation and Selection

In the first subphase that is in key selection phase a 256-bit key is chosen randomly to do the further encryption. This 256-bit key is then divided into four 64bit blocks which will be acted as subkeys at the time of encryption. Each block of subkey is labeled with the DNA base namely A,T,C,G.Then randomly select any combination of these four bases out of the possible 24 combinations without repetition, such as A,C,T,G ; G,A,C,T etc. This key is acted as the round 1 key. Then the key order of DNA bases is right shifted to 1 block that is round 1 key is right shifted to 64 bits. and this process will be continued for 4 times and as a result 4 keys are generated by shifting of 1 block(64 bit) in each round. For example if the randomly chosen key combination is ACGT where each base represents each block of 64 bit. So in the second round the key will be TACG since a block is shifted right, in third and fourth round it will be GTAC and CGTA respectively. Here a single was only chosen and the rest of the four keys were induced in each round from its previous one. In Figure 1 the key generation approach for four consecutive rounds used in the message encryption phase are depicted where the randomly chosen 256 bit key is splitted into four 64 bit subparts and in each round every 64 bit subpart is right shifted to one block and as a result a different key is generated every time in each round. So in this way 4 different keys are generated from a single randomly selected key. In this Figure 1 the dotted arrows signify each right shift operation for each block.

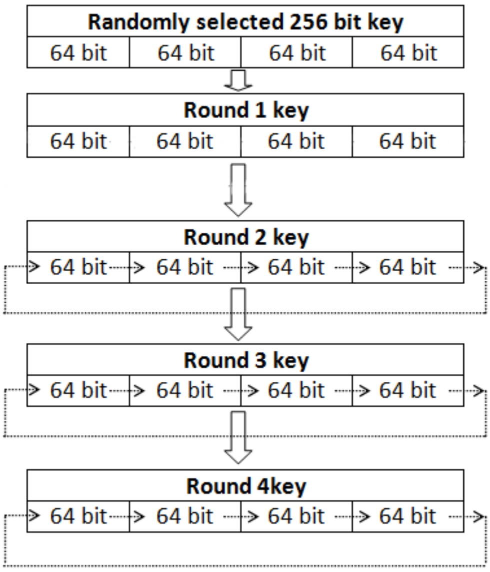

**Figure 1: 4 Round Key Generation Operation**

For Example, Let, randomly selected 256-bit key, K='1000 1010 0011 0001 1100 0100 1011 0001 0111 0000 0010 0100 1111 1100 1111 0011 1110 1110 1011 1011 0100 1011 1111 0000 1011 1100 1101 1011 0011 1010 0001 1110 0001 1000 0011 0110 1010 1111 1000 1110 0001 0000 1001 1010 1001 1001 0010 0001 0000 1101 1010 1110 1000 1111 1000 1111 1101 1011 0011 1010 1111 0101 1111 1110'

Read the 'K' and generate four sub-parts of 64-bit each. Label the subparts with DNA bases (A, T, C, G) as follows:

A='10001010001100011100010010110001011100000001001001111110011110011'

T='11101110101110110100101111111000010111100110110110011101000011110

C='000110000001101101010111111000111000010000100110101001100100100001'

G='00001101101011101000111111000111111011011001110101111010111111110'

Let, randomly selected DNA sequence with 4 DNA bases be 'TGCA' then Round 1 key, K1=TGCA, Round 2 key, K2=ATGC, Round 3 key, K3=CATG, Round 4 key, K4=GCAT.

### Message Encryption

In message encryption phase, the byte values are extracted from the input file or message. The encryption process is works on the unsigned byte values of the input file or text called plaintext. Then these byte values will be transformed into 8-bit binary. Plaintext is then divided into 256 bit blocks; each block of plaintext will go through the encryption process. This phase, the overall encryption is done in 4 consecutive rounds. For each round previously generated keys Key1, Key2, Key3 and Key4 are used to encrypt the plain text. Now each 256bit blocks of plaintext is divided into four 64 bit blocks. After that these 64 bit blocks are further subdivided into 2 parts of 32 bit. Similarly the every 64 bit part of each round selected keys are further subdivided into 2 parts of 32 bit. Then perform an EX-OR operation between the plain text and the selected round encryption key. Rather than directly EX-OR, this EX-OR operation is performed in a different manner where every 32 bit subpart in every 64 bit part of plain text is EX-ORed to the diagonal 32 bit subpart in corresponding 64 bit round key and store it as a new subpart of the intermediate ciphertext. Similarly every part of both plain text and selected round key is EX-ORed and generate a

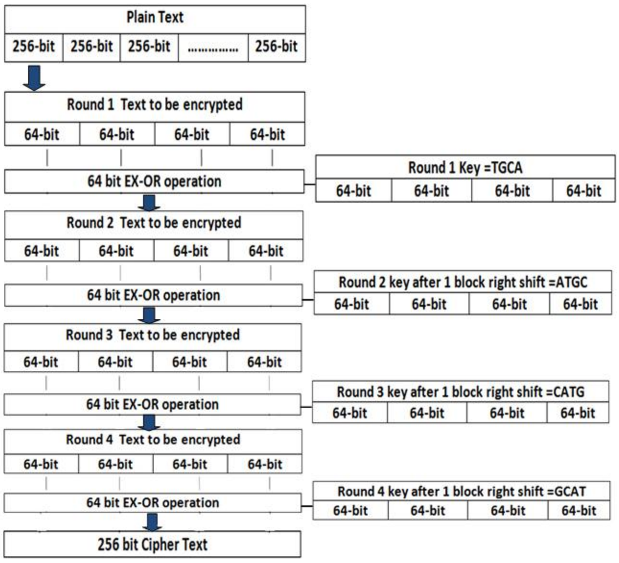

**Figure 2: Encryption Operation for 4 Rounds**

a new 256 bit intermediate cipher text and it will be the input text for the next round. This encryption process is continued upto 4 rounds of encryption using 4 distinct generated keys. The schematic diagram of encryption phase for 4 consecutive rounds is given in Figure 2. The EX-OR operation among the 32bit subparts of both the plain text/intermediate cipher text and the selected round encryption key is depicted in Figure 3.

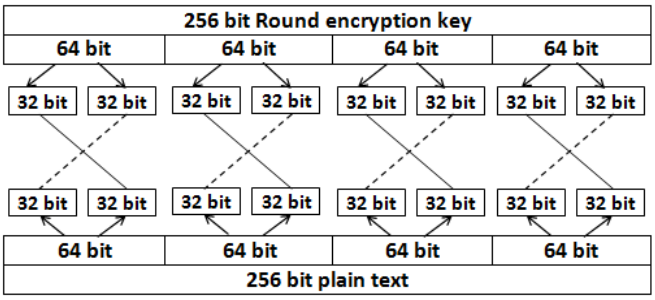

**Figure 3: Proposed EX-OR Operation**

### Algorithmic Steps

**Input:** Input File; Round Keys.

**Output:** Cipher text

**Step 1:** Read the byte values from the input file called plaintext and transform each byte value into 8-bit binary representation.

**Step 2:** Make 256-bit plaintext blocks from the binary representation.

**Step 3:** Repeat step 4 and 11 for each block of plaintext.

**Step 4:** Split the 256-bit block into four 64-bit blocks, namely P1, P2, P3, P4.

**Step 5:** Subdivide each 64 bit Plain text parts into two 32 bit parts, namely $P1_L$, $P1_R$, $P2_L$, $P2_R$, $P3_L$, $P3_R$, $P4_L$, $P4_R$

**Step 6:** Repeat step 7 and 10 for each $Key_i$, where $1 \leq i \leq 4$.

**Step 7:** Read the round encryption key and split into 64 bit parts, namely K1, K2, K3, and K4.

**Step 8:** Subdivide each 64 bit round key parts into two 32 bit parts, namely $K1_L$, $K1_R$, $K2_L$, $K2_R$, $K3_L$, $K3_R$, $K4_L$, $K4_R$

**Step 9:** Compute four 64 bit parts of the Intermediate Cipher Text and store into 4 temporary variables:

$$temp1= Concate\ [(P1_L \oplus K1_R), (P1_R \oplus K1_L)]$$

$$temp2= Concate\ [(P2_L \oplus K2_R), (P2_R \oplus K2_L)]$$

$$temp3= Concate\ [(P3_L \oplus K3_R), (P3_R \oplus K3_L)]$$

$$temp4= Concate\ [(P4_L \oplus K4_R), (P4_R \oplus K4_L)]$$

**Step 10:** Combine all 64-bit cipher blocks to form 256-bit Intermediate cipher text block:

$$ICT = Concate\ (temp1, temp2, temp3, temp4)$$

**Step 11:** Input ICT as input for the next round as plaintext.

**Step 12:** Compute result of round 4 and termed as cipher text CT.

**Step 13:** Club together all the 256-bit cipher text blocks.

### DNA Encoding

In the DNA encoding method of the DNA cryptography, the final cipher text is converted into the DNA form of the data. The coding patterns for encoding the 4 nucleotide bases A, T, C, G is by means of 2-bit binary: 00, 01, 10, and 11 respectively. To enhance the basic DNA characteristics the complementary rule is modified where the complementary rules must follow the rules: $x \neq c(x) \neq c\ (c(x)) \neq c(c(c(x)))$ and $x= c(c(c(c(x))))$.So here the complementary rule (A to T); (T to C); (C to G); (G to A) is used. Here in the proposed approach after getting the cipher CT in the previous phase, it will be transformed into DNA sequence form and compute the complement of the computed DNA sequence and after that a fixed number of bits called primers are to be added at both of the end of the coded DNA. From a publicly available database a DNA sequence of actual plain text size (in bytes) is chosen and split it into two equal parts and treats them as two primers. Now the CT is mapped into a randomly selected array of 16 characters out of 256 ASCII symbols using a hash mapping technique.

Let, the randomly selected hash array bé, H [16] =A, K, Z, S, J, B, T, M, L, F, P, C, R, Y, Q, O.

**Table 1: Hash Mapping Array with DNA Sequence**

| AA | AT | AC | AG | TA | TT | TC | TG | CA | CT | CC | CG | GA | GT | GC | GG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | K | Z | S | J | B | T | M | L | F | P | C | R | Y | Q | O |

Each combination of 2 DNA base sequences is mapped into each corresponding index values of the hash array (H) as per Table 1. As a result the final cipher text FCT is generated.

### Ensuring Integrity

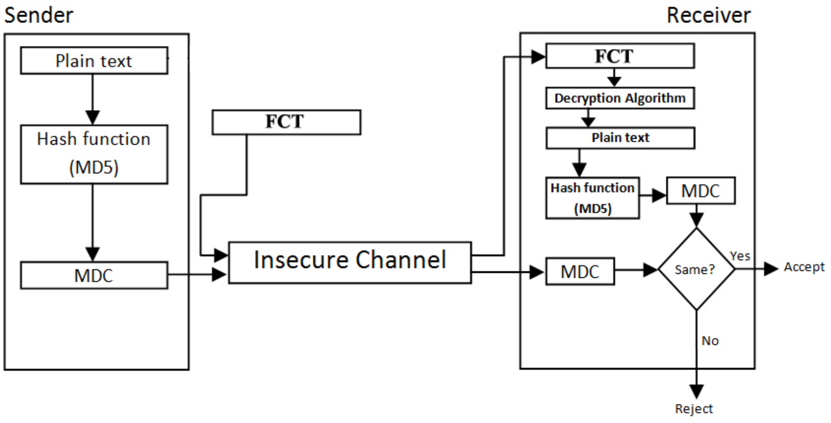

**Figure 4: Message Transmission Ensuring Integrity**

Now to protect the data from modification, insertion, deletion etc we have to ensure the data integrity. For this, the Message detection code (MDC) is followed, which ensures the message integrity. In this phase the sender uses the MD5 as the hash function to create a MDC send it along with the final cipher text FCT to the receiver over an insecure channel. At the receiver side the received FCT is decrypted by following the reverse process of the encryption process used the sender side and the plain text is retrieved. Then the receiver creates a new MDC from the retrieved plain text using the same hash function and the hash key and compares the received MDC and the new MDC. If both of them are same, then it ensures that the message has not been changed. The transmission process with ensuring integrity is depicted in the Figure 4. The schematic diagram of the overall proposed message encoding method explained till now is given in Figure 5.

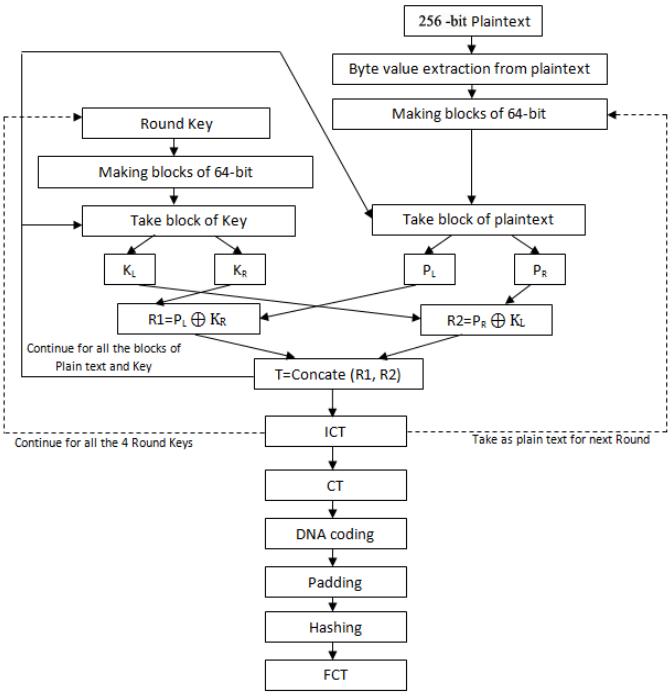

**Figure 5: Schematic Diagram of the Overall Message Encoding Method**

### Decoding at Receiver Side

At the receiver end the randomly selected hash function and the used DNA sequence are available. Receiver can compute the CT from FCT using hash mapping function and binary coding scheme. Remove the extra coding and extract the byte value of the DNA sequence by using the same substitution method as sender had used. Make the cipher text into 64-bit blocks. As the information about the randomly selected Key at the sender end is with the receiver, receiver can compute all the 4 round keys as receiver end has the information about the matrix columns selection for using as keys.

Perform round 1 decryption by K4 and the 256 bit cipher text, which produce intermediate 256 bit blocks of plain text. Then perform round 2 with K3 and intermediate blocks of plain text, which produce next intermediate 256 bit plain text. Then perform same for round 3 with K2 and intermediate plain text, which produce intermediate 256 bit blocks of plain text. At last perform the same for round 4 with K1 and intermediate plain text computed on round 3 and the result of round 4 is the unsigned byte values of the file that has sent by the sender. Write these byte values into a file and save it with its file format. The schematic diagram of the overall proposed message decoding method at the receiver side is given in Figure 6.

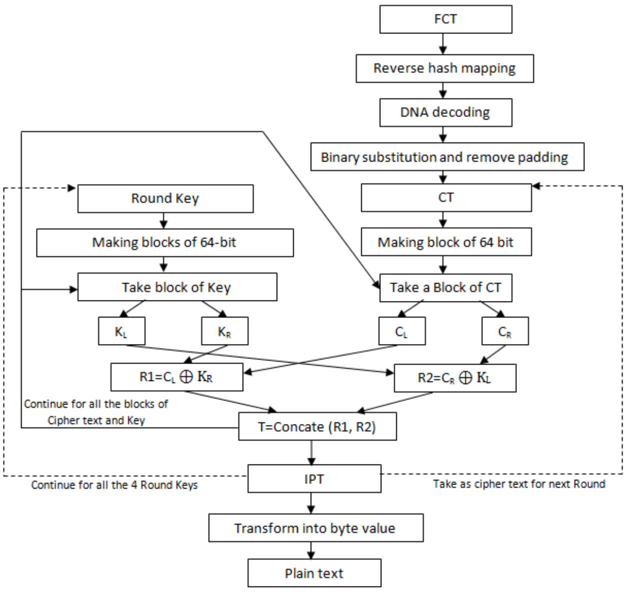

**Figure 6: Schematic Diagram of Receiver Side Message Decoding Method**

## SECURITY AND KEY ANALYSIS

In round operation of encoding 256 bit randomly chosen key is used. Thus if an attacker tries to apply Brute-Force attack on the ciphertext then there are,

$2^{256} = 1.157921 \times 10^{77}$ numbers of possible combinations of keys for cryptanalysis.

In the hash function 16 characters are randomly chosen from 256 ASCII characters. Thus there are,

${}^{256}C_{16} = (256!) / (16! \times (256\text{-}16)!) = 1.00788 \times 10^{25}$ numbers of combinations of keys.

Thus in total for Brute-force attack there are,

$2^{256} \times {}^{256}C_{16} = 1.16704\times 10^{102}$ numbers of possible keys.

Moreover, there are about 163 million DNA sequences available publicly. Thus, probability of an attacker to crack the primer padding in DNA encoding process and make a successful guess on the reference DNA sequence used in DNA encoding step is $1 / (1.6 \times 10^{8})$.

Therefore it could be concluded from the above large key size that it is almost impossible to cryptanalysis and predict the plaintext or the message that is being sent by the sender.

## SECURITY AND KEY ANALYSIS

The algorithm is tested in a system with following configuration:

**Operating System :** Windows 7 (64-bit)

**Platform:** JAVA

**Version:** JDK 7.2

**Processor:** Intel Core i5-2450M, 2.5 GHz

**Primary Memory:** 4 GB DDR3 RAM

The following Table 2 represents the data sets that are obtained during the testing and analysis of the proposed algorithm.

**Table 2: Test Data Sets**

| File Type | File Size (in KB) | Cipher Size (in KB) | Encryption Time (in ms) | Decryption Time (in ms) |
|---|---|---|---|---|
| .doc | 147 | 588 | 6833 | 4846 |
| .pdf | 384 | 1536 | 11466 | 8050 |
| .jpg | 768 | 3072 | 39431 | 22371 |
| .mp3 | 3105 | 12420 | 62712 | 28314 |
| .flv | 4028 | 16112 | 78717 | 34804 |

## CONCLUSIONS

In this approach the secret information has been hidden in more depth so that it will be almost impossible to an attacker to know about the message. Here a long size 256-bit key is used in different way each time in 4 rounds to encrypt the message. Moreover it is almost impossible to an intruder to predict DNA sequence. The message encryption approach is also better than the available cryptographic algorithms based on the DNA due to using some complex computations performed on the data. The main strength of this proposed method is that beside the security and confidentiality it also provides integrity in a greater extent. Thus it will very much difficult for the intruders to apply different cryptanalysis on the cipher text.

## REFERENCES

1. H. Z. Hsu and R.C.T.Lee. (2006). DNA Based Encryption Methods. The 23rd Workshop on Combinatorial Mathematics and Computation Theory, National Chi Nan University Puli, Nantou Hsies. Taiwan 545.
2. Amal Khalifa and Ahmed Atito. (2012). High-Capacity DNA-based Steganography. In the 8th International Conference and informatics and Systems (INFOS2012).IEEE.
3. Mohammad Reza Abbasy, Pourya Nikfard, Ali Ordi, Mohammad Reza Najaf Torkaman. (2012). DNA Base Data Hiding Algorithm. In International Journal on New Computer Architectures and their Applications.
4. Sabari Pramanik, Sanjit Kumar Setua. (2012). DNA Cryptography. In ICECE, 2012, pp. 551-554. IEEE. doi:10.1109/ICECE.2012.6471609.
5. Suman Chakraborty, Sudipta Roy, Prof. Samir K. Bandyopadhyay. (2012). Image Steganography Using DNA Sequence and Sudoku Solution Matrix. In International Journal of Advanced Research in Computer Science and Software Engineering.
6. Nirmalya Kar, Atanu Majumder, Ashim Saha, Anupam Jamatia, Kunal Chakma, Dr. Mukul Chandra Pal. (2013). In MobileHealth'13.Bangalore, India.ACM.
7. Bibhash Roy, Atanu Majumder. (2012). An Improved Concept of Cryptography Based on DNA Sequencing. In International Journal of Electronics Communication and Computer Engineering. Vol. 3, Issue-6.

8. Bibhash Roy, Gautam Rakshit, Pratim Singha, Atanu Majumder, Debabrata Datta. (2011). An improved Symmetric Key Cryptography with DNA Based Strong Cipher. In ICDeCom-2011, BIT Mesra, Ranchi, Jarkhand, India.

9. Xing Wang, Qiang Zhang. (2009). DNA computing-based cryptography. In Proc. of the 2009 IEEE International Conference, ISBN: 978-1-4244-3867-9/09.

10. Guangzhao Cui, Limin Qin, Yanfeng Wang, Xuncai Zhang. (2008). An Encryption Scheme Using DNA Technology", In BICTA 2008, pp. 37-42.IEEE. doi:10.1109/BICTA.2008.4656701.

11. Behrouz A.Forouzen, Debdeep Mukhopadhyay. Cryptography and Network Security. 2nd Edition, Tata McGraw Hill Education Pvt.Ltd.